Assessment of the dynamics of microparasite infections in genetically homogeneous and heterogeneous populations using a stochastic epidemic model

Assessment of the dynamics of microparasite infections in genetically homogeneous and heterogeneous populations using a stochastic epidemic model

M. Nath, J. A. Woolliams and S. C. Bishop

J ANIM SCI 2008, 86:1747-1757.
doi: 10.2527/jas.2007-0615 originally published online April 11, 2008

Assessment of the dynamics of microparasite infections in genetically homogeneous and heterogeneous populations using a stochastic epidemic model[1]

M. Nath,[2,3] J. A. Woolliams, and S. C. Bishop

Roslin Institute and Royal (Dick) School of Veterinary Studies, Roslin BioCentre, Midlothian, EH25 9PS, United Kingdom

ABSTRACT: The aim of this paper was to explore the effect of genetic heterogeneity in host resistance to infection on the population-level risks and outcomes of epidemics. This was done using a stochastic epidemiological model in which the model parameters were assumed to be genetically controlled traits of the host. A finite locus model was explored, with a gene controlling the transmission coefficient (i.e., host susceptibility to infection) and a gene controlling the recovery period. Both genes were simulated to have 2 alleles with underlying additive or dominance inheritance and an independent assortment of alleles. The model was parameterized for a viral pig disease (transmissible gastroenteritis), and complete homogeneous mixing among genotypes was assumed. Mean population genotype dramatically affected epidemic outcomes, and subtle effects of heterogeneity on epidemic properties were also observed. Genetic variation in the transmission coefficient led to probabilities of epidemics occurring that were slightly greater than expected, but genetic variation in the recovery rate had no such effect. Epidemics were generally less severe in genetically heterogeneous populations than expected from the constituent subpopulations. Furthermore, the genotype of the initial infected animal had a marked effect on epidemic probabilities, particularly when genetic variation was for recovery rate. The results of this model provide useful information to determine the optimum population structures and to exploit genetic variation in resistance to infection. Applications of the proposed model in genetically heterogeneous populations for identifying practical disease management strategies are also discussed.

Key words: animal health, disease resistance, epidemiology, genetics, infection transmission

J. Anim. Sci. 2008. 86:1747–1757
doi:10.2527/jas.2007-0615

INTRODUCTION

Exploiting genetic variation in resistance to infection may be an economical and environment-friendly approach to control livestock diseases. Genetic differences in host resistance to infection have been documented in all major domestic species (Axford et al., 2000), and large research efforts aim to detect genes explaining this variation. For example, using a combination of fine-mapping and functional genomics, Cheng (2005) reported several genes involved in host-pathogen interactions for Marek's disease in poultry. Many such genes are likely to be identified that contribute to host disease resistance. This leads to the question: how do we choose among various genes when breeding for disease resistance, and what will be the effect of utilizing different genes that influence different aspects of the host-pathogen interaction?

To answer such questions, we need epidemiological models that capture genetic heterogeneity in disease resistance. Published models with host heterogeneity (Adler, 1992; Dushoff and Levin, 1995; Yates et al., 2006) consider nongenetic factors such as behavior and mixing pattern. Moreover, these models tend to be deterministic and as such do not capture stochastic attributes of the transmission of infection (Renshaw, 1991). Stochastic epidemic models (Bouma et al., 1995; Innocent et al., 1997; Stark et al., 2000) also did not consider host genetics. We have previously used stochastic genetic epidemiological models to quantify the effects of altering the mean host resistance genotype (MacKenzie and Bishop, 2001; Nath et al., 2004) and

[1]The authors thank the Biotechnology and Biological Sciences Research Council for financial support.
[2]Corresponding author: mintu@bioss.ac.uk
[3]Present address: Biomathematics and Statistics Scotland, JCMB, The King's Building, Edinburgh, EH9 3JZ, United Kingdom.
Received September 27, 2007.
Accepted March 25, 2008.

predict future disease risks (Springbett et al., 2003). The aim of the present paper is to link these concepts and investigate the properties of epidemics in populations containing a mixture of relatively susceptible and resistant genotypes at the individual locus level, using equivalent homogeneous populations as a reference point, as well as to explore strategies relevant to genetic management of disease.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from a computer-generated data set.

Overview of Methodology

This paper explores the properties of epidemics, in closed homogeneous and heterogeneous populations, caused by microparasite infections. To achieve this, our approach was as follows. We developed a stochastic model describing the transmission of infection and the severity of subsequent epidemics. An important livestock disease was identified for which parameter estimates were available from the published literature. Hypothetical allelic effects for genes controlling rate-limiting parameters were defined, and the model was developed assuming both additive and dominance modes of inheritance. Finally, the model was applied in both homogeneous and heterogeneous populations, and epidemiological outputs were obtained.

For ease of interpretation, we assumed that a single locus had a large influence on aspects of either resistance or susceptibility to infection. Although this is a simplifying assumption, there are several cases where it is realistic [e.g., major histocompatibility complex effects for resistance to Marek's disease; the PrP locus for resistance to scrapie in sheep; major genes for *Escherichia coli* resistance in pigs (Bishop, 2005)], and it does not preclude the influence of other major loci or even background polygenic effects. Further, this parameterization represents cases for which selection for disease resistance may be based initially on a single locus.

Infection Process and Parameter Definition

The transmission of infection during microparasitic epidemics may be described by so-called compartmental models (Anderson and May, 1992), in which animals move from one state to another, defined according to the events that follow as a result of infection. These are as follows. First, a susceptible individual may be infected by an infectious individual. The transmission coefficient, ß, denotes the rate at which infectious individuals transmit infection to susceptible animals and is the expected number of new infections per infectious animal per susceptible animal per day. In the context of disease genetics, the parameter ß depends on the infectivity genotypes of infectious animals and the susceptibility genotypes of susceptible animals. In this paper, we modeled ß as the susceptibility genotype, and hence, this is a parameter associated with the likelihood of susceptible individuals becoming infected. The second model parameter for which we modeled genetic variation, recovery period, was associated with attributes of the infected individuals (as described below).

After infection, a latent period may ensue, which is the period during which the individual is infected but noninfectious (i.e., not yet capable of transmitting the infection), because the abundance of the infectious agent is low. The parameter, σ, is the inverse of the expected latent period. The next phase is the infectious period, when the abundance of pathogens is high and the individual is capable of infecting other susceptible individuals. After the infectious phase, host recovery may occur when abundance of pathogens decreases to zero or to a low level. The parameter, γ, is the reciprocal of the infectious period (or recovery period) and is the expected number of recoveries per infected animal per day. Infected individuals also may die as a result of infection. The mortality rate, ε, describes the disease-dependent mortality (i.e., the expected proportion of individuals dying per day among the infected individuals). Finally, immunity in the recovered individuals may not be lifelong and may persist only for a period of time, after which they become susceptible again. The parameter, ω, is represented by the reciprocal of the expected period between the time of recovery and the time when recovered individuals again become susceptible (loss of immunity period). Here it is assumed that $\omega > 0$ (i.e., that immunity is not lifelong).

Epidemic Model

The stochastic epidemic model simulates a series of random events in time, with the probability of specific events defined by the parameters of the model. The possible event types for this model are as follows: (1) a susceptible animal becomes latently infected, (2) a latently infected animal becomes infectious, (3) an infected animal recovers or dies as a result of infection, and (4) an animal that has recovered loses immunity and becomes immunologically susceptible again. This model can be abbreviated as ***SLIRDS*** (for susceptible-latent-infected-recovered/died-susceptible).

The model also simulates the interevent time. For a population of N animals with S susceptible animals, L animals in the latent class, I infected animals, Q recovered animals, and n genotypes, the epidemic interevent time has a mean of:

$$1\Big/\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_j c_{ji} I_i S_j + \sum_{i=1}^{n}\sigma_i L_i + \sum_{i=1}^{n}\gamma_i I_i + \sum_{i=1}^{n}\varepsilon_i I_i + \sum_{i=1}^{n}\omega_i Q_i\right), \quad [1]$$

where c_{ji} = the contact rate between a susceptible animal of genotype j and an infectious animal of genotype

i (assumed to be 1 in this paper), and the other parameters are as defined above. The interevent time may then be drawn from an exponential distribution as: $-ln(r) \times$ (mean interevent time), where r = a random number in [0, 1].

The next event type is determined by calculating the probability of a specific event in relation to all possible events (***RATE***). Thus, let Eq. [1] be *RATE*. Then, the probabilities of the next event are as follows: (1) the infection of an animal, moving the newly infected animal to the latent class, is $\beta SI \times RATE$; (2) the movement of a latent animal to the infectious class is $\sigma L \times RATE$; (3) the recovery of an infected animal is $\gamma I \times RATE$; (4) the death of an infected animal is $\varepsilon I \times RATE$; and (5) the loss of immunity of a recovered animal is $\omega Q \times RATE$. Thus, the precise event can be determined by sampling from a random uniform distribution.

Population Structure and Parameter Space Investigated

As a case study, we considered transmissible gastroenteritis, a highly contagious enteric viral disease of swine. Estimates of the benchmark parameters describing transmissible gastroenteritis have been published for growing pigs (Hone, 1994), and the values used were as follows: transmission coefficient (0.00007), latent period (2 d), recovery period (20 d), mortality rate (0.02), and days before loss of immunity (180 d).

We considered 10,000 intensively housed pigs of the same age and physiological status. The population was assumed to be closed, with no introductions of animals during the epidemic period, no disease-independent mortality, and no external intervention (such as medication, vaccination, isolation, culling, etc.) on the course of the epidemic.

We modeled genetic variation in resistance to infection as being influenced by 2 unlinked loci, namely B and R, which represent the parameters for the transmission coefficient (susceptibility of susceptible individuals) and the recovery period of infected individuals, respectively. The favorable allele for the B locus is *B,* and the unfavorable allele is *b*. Similarly, the favorable and unfavorable alleles for the R locus are *R* and *r*, respectively. Due to independent assortment, a total of 9 genotypic classes are expected. We considered 2 modes of inheritance, additive and dominance. Under the dominance model, the unfavorable allele was dominant over the favorable allele (namely, the *b* allele was dominant over the *B* allele and the *r* allele was dominant over the *R* allele).

For additive models, we assumed the heterozygote to have the benchmark parameters, with homozygotes being 0.5 and 1.5 times the benchmark. Thus, the genotypic values for the B locus were 0.000035, 0.000070, and 0.000105 for *BB*, *Bb*, and *bb*, respectively. Similarly, the genotypic value of *Rr* was 20 d ($\gamma = 0.05$), and the values for *RR* and *rr* were 13.3 d ($\gamma = 0.075$) and 40 d ($\gamma = 0.025$), respectively. In the case of dominance effects, genotypic values of the heterozygotes and the homozygotes for unfavorable alleles were the same ($\beta = 0.000070$ and $\gamma = 0.05$). These parameterizations do not rule out the influence of other loci, but for simplicity, other gene effects are not modeled in this paper. To make the comparison of different models easier, the latent period and the period of immunity after recovery were set to the benchmark parameter values for all individuals (namely, a latent period of 2 d and 180 d before loss of immunity).

Capturing the Output

Probability of an Epidemic. The probability that an epidemic will occur, given the presence of an infected individual, was determined by the proportion of simulations that resulted in an epidemic. If the infected individual, introduced in the population at the beginning of the epidemic, recovers or dies without causing any further infections, then it is considered to be no epidemic. Otherwise, epidemics may be split into those that are major and those that are minor epidemics. Minor epidemics are those that die out due to chance events without intervention (Bishop and MacKenzie, 2003). For practical purposes, thresholds may be set to distinguish between major and minor epidemics. In these simulations, if more than 10% of the individuals of the population became infected, the epidemic was deemed as being major. Otherwise, it was considered to be minor. The choice of 10% of animals infected as the threshold was made on the grounds that epidemics that died out quickly generally resulted in fewer than 10% of the population becoming infected, for the parameter combinations investigated.

Basic Reproductive Ratio. The basic reproductive ratio, R_0, is a dimensionless parameter that encapsulates the biological details of different transmission mechanisms. For microparasites, R_0 is the expected number of infections produced by the introduction of an infected individual into an otherwise completely susceptible population, during the course of its infectious period (Diekmann et al., 1990). For the susceptible-infected-recovered (***SIR***) model, the probability of no epidemic, p, is $1/(R_0 + 1)$ (Bishop and MacKenzie, 2003). Hence, R_0 was approximated as $(1/p) - 1.0$. Because the probability of no epidemic is defined equivalently for the *SIR* and *SLIRDS* models, because it depends only on whether the index case infects a further animal before the index case recovers or dies, the subsequent definition of R_0 from epidemic probabilities is the same for both models. The SD of R_0 was estimated from 1,000 bootstrapped samples, in which for each bootstrapped sample, outputs of all 10,000 replicates (see below) were sampled with replacement.

Epidemic Severity. Epidemic severity was defined in terms of the maximum proportion of animals infected at any one time during the epidemic, y_{max}, and the time of its occurrence, t_{max}, was also noted. The estimate of y_{max} was averaged for replications when

Table 1. Combinations of genotypes investigated in heterogeneous populations of 10,000 animals and the respective numbers of animals of each genotype

Genotype	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
BBRR	625	624	2,500	2,500	0	0	2,500	0	0
BBRr	1,250	938	0	5,000	0	0	0	2,500	0
BBrr	625	938	0	2,500	0	0	0	0	2,500
BbRR	1,250	938	0	0	2,500	0	5,000	0	0
BbRr	2,500	2,500	5,000	0	5,000	0	0	5,000	0
Bbrr	1,250	1,562	0	0	2,500	0	0	0	5,000
bbRR	625	938	0	0	0	2,500	2,500	0	0
bbRr	1,250	1,562	0	0	0	5,000	0	2,500	0
bbrr	625	0	2,500	0	0	2,500	0	0	2,500

a major epidemic occurred. Simulations of major epidemics (i.e., those with >10% of the animals infected), were continued for 6 in silico months to capture the estimates and SD of y_{max} and t_{max}.

Simulation Process

Each simulation began by introducing a single infected individual (the index case) into the farm. The initial interevent time was calculated, the first event was determined, and the epidemic commenced. The epidemic continued until no infected animals remained in the population (i.e., the epidemic had ceased) or 6 in silico months had elapsed.

Simulations were run in genetically homogeneous and heterogeneous populations. The simulations for the homogenous population considered each of 9 genotypes individually. The heterogeneous population comprised combinations of 3, 8, and 9 genotypes. The allelic frequency was set to 0.5 for both loci, and corresponding genotypic frequencies were calculated in all cases so that the overall allele frequencies remained unaltered in the heterogeneous population (cases 1, 2, and 3). Cases were also considered with fixed B genotypes and variable R genotypes (cases 4, 5, and 6) and fixed R genotypes and variable B genotypes (cases 7, 8, and 9). All cases and their corresponding frequencies are presented in Table 1.

A total of 10,000 replicates were run for each scenario, and the output parameters were obtained by pooling over the 10,000 replicates. Additionally, expected values of the epidemiological parameters in the heterogeneous population were estimated from the parameter values of the homogeneous populations. To capture outcomes due to differences in the index case genotype in the heterogeneous population, each genotype in turn was considered as the genotype of the index case, with the number of replicates proportional to the proportion of the population that the genotype comprised.

Individual-Based Model

Finally, an individual-based epidemic model was also considered to estimate the number of infections caused by each individual of a particular genotype along with the index case and also to obtain the number of days that an infected individual remained in the population before recovering. Because the aim was to identify the genetic heterogeneity in terms of ß and γ, the individual-based model was explored considering $\mu = 0$ and $\omega = 0$. Hence, this was a *SIR* model. A total of 1,000 individuals were considered in the individual-based model. Because the parameter ß is proportional to population size, a value of ß = 0.0007 was used for the homogeneous population. The recovery rate parameter for the homogeneous population was the same as the benchmark parameter value of the *SLIRDS* model (i.e., γ = 0.05).

To create a heterogeneous, individual-based, *SIR* epidemic model, the population was defined to comprise individuals of 3 genotypes (*BBRR*, *BbRr*, and *bbrr*) having variable transmission coefficients (0.00035, 0.00070, and 0.00100) and recovery rates (0.075, 0.050, and 0.025), with alleles in Hardy-Weinberg equilibrium. In all scenarios, a single index case of a specific genotype was introduced into the population, and 5,000 replicates were run. The simulation strategies and model outputs were similar to those for the *SLIRDS* model. The individual-based model aided in capturing the detailed behavior of each individual in the population and its influence on the epidemic dynamics.

RESULTS

Homogeneous Population

The estimates of epidemic probabilities for the homogeneous population and different epidemiological parameters are presented in Table 2. These are the results against which the heterogeneous population results are to be compared. From the estimates of the different epidemic parameters, it is evident that under the assumptions of additive and dominance effects of alleles, the results followed a more or less similar pattern. In general, the results demonstrated that mean population genotype can have dramatic effects on epidemic outcomes, and most results may be predicted from input parameters using standard epidemic theory. It is noteworthy that the time of the maximum epidemic severity (t_{max}) was critically affected by the transmission coefficient rather than by the recovery rate, and it is therefore not strictly a function of R_0. On

Table 2. Epidemic properties in different homogeneous populations comprising a specific genotype under additive and dominance models[1]

	Epidemic probabilities			R_0 properties		Epidemic severity		Epidemic timing	
Genotype	No epidemic	Minor epidemic	Major epidemic	R_0	SD	y_{max}	SD	t_{max}	SD
Additive model									
BBRR	0.32	0.07	0.61	2.09	0.05	0.31	0.01	60.9	5.86
BBRr	0.25	0.04	0.71	2.94	0.07	0.42	0.01	57.5	5.21
BBrr	0.17	0.02	0.81	4.76	0.12	0.55	0.01	55.2	4.68
BbRR	0.22	0.03	0.75	3.49	0.09	0.48	0.01	33.8	2.64
BbRr	0.17	0.02	0.81	4.90	0.13	0.57	0.01	33.2	2.49
Bbrr	0.11	0.01	0.88	7.86	0.25	0.68	0.01	33.0	2.37
bbRR	0.18	0.02	0.80	4.61	0.13	0.55	0.01	25.3	1.77
bbRr	0.13	0.01	0.86	6.51	0.18	0.63	0.01	25.2	1.71
bbrr	0.09	0.00	0.91	10.2	0.35	0.72	0.01	25.4	1.66
Dominance model									
BBRR	0.32	0.07	0.61	2.09	0.04	0.31	0.01	60.9	5.86
D1	0.25	0.04	0.71	2.94	0.07	0.42	0.01	57.5	5.21
D2	0.22	0.03	0.75	3.50	0.08	0.48	0.01	33.8	2.64
D3	0.17	0.02	0.81	4.91	0.14	0.57	0.01	33.2	2.49

[1]Similar epidemic properties observed for 2 or more genotypes under a dominance model are presented together as *D1*, *D2*, and *D3*. Different genotypes under these are as follows: *D1: BBRr* and *BBrr*; *D2: BbRR* and *bbRR*; *D3: BbRr, Bbrr, bbRr*, and *bbrr*.

the other hand, the estimates of the maximum proportion of population infected (y_{max}) varied with variation in both the transmission coefficient and the recovery rate, and the estimates of R_0 changed nonlinearly with estimates of y_{max}. The bootstrapped SD of R_0 were very small, confirming the precision of the estimates in a simulation of this magnitude.

To explore the variability that may be expected from different epidemics within the same population, we captured the frequency of infections that were directly caused by introduction of a single index case in the population of susceptible individuals, using a simple *SIR* model. The number of cases when the index individual recovered without causing any subsequent infections was noted, and the estimated R_0 value of the population was calculated using the formula $R_0 = 1/p - 1.0$. The estimate of R_0 obtained stochastically (14.4) was close to the deterministic expectation (i.e., $\beta N/\gamma = 14.0$; data not shown).

Heterogeneous Population

The estimates of epidemic probabilities and R_0 properties for the heterogeneous population are presented in Table 3, and indicators of epidemic severity are shown in Table 4. In all cases, the empirical SD of the observed parameters and expected values of parameters (calculated as weighted averages of estimates across homogeneous populations) are presented. The results obtained from heterogeneous populations under additive and dominance models showed similar trends.

The epidemic dynamics in heterogeneous populations showed some interesting properties under the present conditions of modeling. In terms of epidemic probabilities, when the populations consisted of fixed transmission coefficient genotypes but variable recovery rates (e.g., cases 4, 5, 6), the observed probabilities of no epidemic were more than the expected values. These were reflected by lower estimates of R_0 in the heterogeneous population than that expected from the weighted average of constituent homogenous populations. On the other hand, where the recovery rate genotypes were constant but the transmission coefficients varied (e.g., cases 7, 8, 9), observed probabilities of no epidemic were generally slightly less than the expected probabilities, which meant that the observed values of R_0 tended to be greater than the expected values (Table 3). Further, in all cases, the expected major epidemic probabilities were either less than or equal to the observed major epidemic probabilities. Hence, the expected values of major epidemic probabilities slightly underestimate the actual major epidemic scenario.

The estimates of epidemic severities (y_{max}) in heterogeneous populations were almost unaffected by the genetic heterogeneity, and all estimates were close to the expected values obtained from the constituent subpopulations. However, the estimates of observed timing of maximum epidemic severity (t_{max}) were always smaller than the expected values (Table 4). As in homogeneous populations, y_{max} varied with changes in the transmission coefficient and the recovery rate, and t_{max} was observed to vary with changes in the population mean transmission coefficient but not with recovery rate.

The effect of genetic heterogeneity, as defined by having highly susceptible animals in the population, on the overall disease dynamics can also be appreciated by comparing populations under cases 1 and 2. The population under case 2 has the same mean genotypic levels for the 2 loci as population 1, but it lacks the most susceptible genotype. From theoretical expec-

Table 3. Observed no, minor, and major epidemic probabilities and observed (Obs.) basic reproductive ratio (R_0) in heterogeneous populations comprising combinations of different genotypes under additive and dominance models, along with expected (Exp.) values obtained from corresponding homogeneous populations

		Epidemic probabilities				R_0		
Case	Genotypes	No	Minor	Major	Exp. major	Obs.	SD	Exp.
Additive model								
1	All 9 genotypes	0.17	0.01	0.82	0.80	4.84	0.13	5.18
2	All but *bbrr*	0.17	0.02	0.81	0.80	4.98	0.13	5.08
3	*BBRR*, *BbRr*, *bbrr*	0.15	0.01	0.84	0.79	5.49	0.15	5.53
4	*BBRR*, *BBRr*, *BBrr*	0.25	0.04	0.71	0.71	3.04	0.07	3.18
5	*BbRR*, *BbRr*, *Bbrr*	0.17	0.01	0.82	0.81	4.92	0.13	5.29
6	*bbRR*, *bbRr*, *bbrr*	0.13	0.01	0.86	0.86	6.59	0.20	6.96
7	*BBRR*, *BbRR*, *bbRR*	0.22	0.03	0.75	0.73	3.60	0.09	3.42
8	*BBRr*, *BbRr*, *bbRr*	0.15	0.01	0.84	0.80	5.77	0.16	4.82
9	*BBrr*, *Bbrr*, *bbrr*	0.11	0.01	0.88	0.87	7.96	0.25	7.67
Dominance model								
1	All 9 genotypes	0.20	0.02	0.78	0.77	4.08	0.11	4.10
2	All but *bbrr*	0.19	0.02	0.79	0.77	4.12	0.10	4.10
3	*BBRR*, *BbRr*, *bbrr*	0.19	0.02	0.79	0.76	4.41	0.11	4.20
4	*BBRR*, *BBRr*, *BBrr*	0.27	0.04	0.69	0.68	2.71	0.06	2.73
5	*BbRR*, *BbRr*, *Bbrr*	0.18	0.02	0.80	0.80	4.52	0.12	4.55
6	*bbRR*, *bbRr*, *bbrr*							
7	*BBRR*, *BbRR*, *bbRR*	0.24	0.03	0.73	0.72	3.18	0.07	3.15
8	*BBRr*, *BbRr*, *bbRr*	0.18	0.02	0.80	0.79	4.69	0.12	4.41
9	*BBrr*, *Bbrr*, *bbrr*							

tations, the estimates of expected R_0 were 5.18 under case 1 compared with 5.08 under case 2, and the results from Tables 3 and 4 indicated that under both additive and dominance models the epidemic severities were similar in both cases. Thus, the presence or absence of small numbers of the most susceptible genotype has had little effect, provided that the overall mean genotypic effects are constant.

The results described above represent the average outcomes for the heterogeneous populations irrespective of the genotype of the index case. However, the index case genotype also has an effect upon the initial events in the epidemic (Table 5). Depending on the susceptibility or resistance ability of the index case genotype, there was a large change in the probability of the major epidemic in the heterogeneous population. For

Table 4. Observed (Obs.) epidemic severity and timing in heterogeneous populations comprising combinations of different genotypes under additive and dominance models, along with expected values (Exp.) obtained from corresponding homogeneous populations

		Epidemic severity			Epidemic timing		
Case	Genotypes	Obs. y_{max}	SD	Exp. y_{max}	Obs. t_{max}	SD	Exp. t_{max}
Additive model							
1	All 9 genotypes	0.55	0.01	0.55	33.6	2.44	37.4
2	All but *bbrr*	0.54	0.01	0.55	33.7	2.43	37.3
3	*BBRR*, *BbRr*, *bbrr*	0.56	0.01	0.54	33.3	2.43	38.2
4	*BBRR*, *BBRr*, *BBrr*	0.43	0.01	0.43	57.6	5.12	57.8
5	*BbRR*, *BbRr*, *Bbrr*	0.57	0.01	0.58	33.2	2.43	33.3
6	*bbRR*, *bbRr*, *bbrr*	0.63	0.01	0.63	25.2	1.70	25.2
7	*BBRR*, *BbRR*, *bbRR*	0.45	0.01	0.46	34.0	2.60	38.4
8	*BBRr*, *BbRr*, *bbRr*	0.54	0.01	0.55	33.6	2.45	37.3
9	*BBrr*, *Bbrr*, *bbrr*	0.65	0.01	0.66	33.6	2.36	36.6
Dominance model							
1	All 9 genotypes	0.51	0.01	0.51	36.9	2.77	39.6
2	All but *bbrr*	0.51	0.01	0.51	36.9	2.77	39.6
3	*BBRR*, *BbRr*, *bbrr*	0.51	0.01	0.51	36.7	2.87	40.1
4	*BBRR*, *BBRr*, *BBrr*	0.39	0.01	0.39	58.3	5.35	58.4
5	*BbRR*, *BbRr*, *Bbrr*	0.55	0.01	0.55	33.3	2.47	33.3
6	*bbRR*, *bbRr*, *bbrr*						
7	*BBRR*, *BbRR*, *bbRR*	0.44	0.01	0.44	37.5	3.04	40.6
8	*BBRr*, *BbRr*, *bbRr*	0.53	0.01	0.53	36.7	2.84	39.3
9	*BBrr*, *Bbrr*, *bbrr*						

Table 5. Effect of the index case genotype on epidemic probabilities in a heterogeneous population of 10,000 animals under additive and dominance models

		Additive model			Dominance model		
Index genotype	Genotype frequency	No epidemic	Minor epidemic	Major epidemic	No epidemic	Minor epidemic	Major epidemic
BBRR	625	0.21	0.01	0.78	0.22	0.04	0.74
BBRr	1,250	0.17	0.01	0.81	0.18	0.02	0.80
BBrr	625	0.14	0.01	0.86	0.20	0.01	0.79
BbRR	1,250	0.22	0.02	0.76	0.26	0.03	0.72
BbRr	2,500	0.17	0.01	0.82	0.19	0.02	0.79
Bbrr	1,250	0.10	0.01	0.89	0.18	0.02	0.80
bbRR	624	0.23	0.02	0.75	0.24	0.02	0.74
bbRr	1,250	0.18	0.01	0.81	0.18	0.02	0.80
bbrr	625	0.12	0.01	0.87	0.16	0.01	0.83

example, when the most resistant (*BBRR*) and most susceptible (*bbrr*) genotypes were considered as index cases, the R_0 values of the whole population increased from 3.74 to 9.20 under an additive model and from 3.37 to 5.67 under a dominance model. Further, it is evident from Table 5 that the effect of the index case depends upon the recovery genotype rather than the transmission coefficient genotype. However, the maximum proportion of infection (y_{max}) and time of occurrence of y_{max} (t_{max}) remained the same when different genotypes were considered as index cases (results not shown). Hence, the parameters describing the subsequent development of the epidemic are unaffected by the index case.

Using an individual-based simple *SIR* epidemic model, we illustrated the epidemic dynamics further in a population comprising individuals of 3 genotypes (*BBRR*, *BbRr*, and *bbrr*) having variable transmission coefficients (0.00035, 0.00070, 0.00100) and recovery rates (0.075, 0.050, 0.025) with alleles in Hardy-Weinberg equilibrium. Results were summarized over all simulations to identify the number of infections caused by the index case, as well as the day when the index case recovered. Figure 1(a) presents the frequencies of infections caused by the index case of each genotype, whereas Figure 1(b) presents the frequencies for day of recovery of the index case of each genotype. Figure 1(a) indicates that the number of times the index case recovered without causing any infections was almost 1.5 and 3.0 times greater for *BBRR* than *BbRr* and *bbrr*, respectively. From both the *SIR* model results and the *SLIRDS* model results (Table 5 for *SLIRDS*), it can be concluded that the differences in frequencies of infections essentially corresponded to the parameter recovery rate of each genotype.

The differences in frequencies of infections were also reflected in the frequencies of day for recovery of different genotypes as displayed in Figure 1(b). The number of times the index case of *BBRR* recovered before the first day of epidemic was 1.6 and 3.2 times more than that for *BbRr* and *bbrr*. On the other hand, the frequencies of 6 or more infections were always greater in *bbrr* than in *BBRR* and *BbRr*. Figure 1(b) also indicated that as the epidemic progressed, less than 0.01% of cases were observed when a *BBRR* index case individual remained infected beyond the 56th day, whereas this was so for the 114th and 136th day when the index case was *BbRr* and *bbrr*. No infected individual of the index case *BBRR* and *BbRr* was observed beyond the 103rd and 139th day, respectively, whereas there were several occasions that the index case of *bbrr* remained infected for the whole duration of the epidemic and continued to infect other susceptible individuals in the population. Hence, index case individuals of the most susceptible genotype remained in the population for longer duration and resulted in a greater number of subsequent infections.

Our simulation results also captured information on individuals of different genotypes, to explore their role in causing subsequent infections. The results are presented graphically in Figure 2(a) and 2(b), respectively, for the frequencies of subsequent infections and day of recovery for individuals of different genotypes. These figures present the results considering only the *BbRr* genotype as the index case. However, outcomes were of a similar pattern when the index cases of other genotypes were considered. The frequencies of no subsequent infection were the greatest in *BBRR* followed by *BbRr* and *bbrr*, whereas frequencies of one or more infections were in reverse order (i.e., greater in *bbrr* followed by *BbRr* and *BBRR*). Similarly, the frequencies for day of recovery indicated that greater numbers of *BBRR* individuals recovered early in the epidemic, whereas *bbrr* individuals remained infected in the population for a prolonged time.

DISCUSSION

Effect of Genetic Heterogeneity

Effects of genetic heterogeneity were observable but subtle. Our results indicated that genetic variation in the recovery rate resulted in probabilities of no epidemic that were slightly greater than that ex-

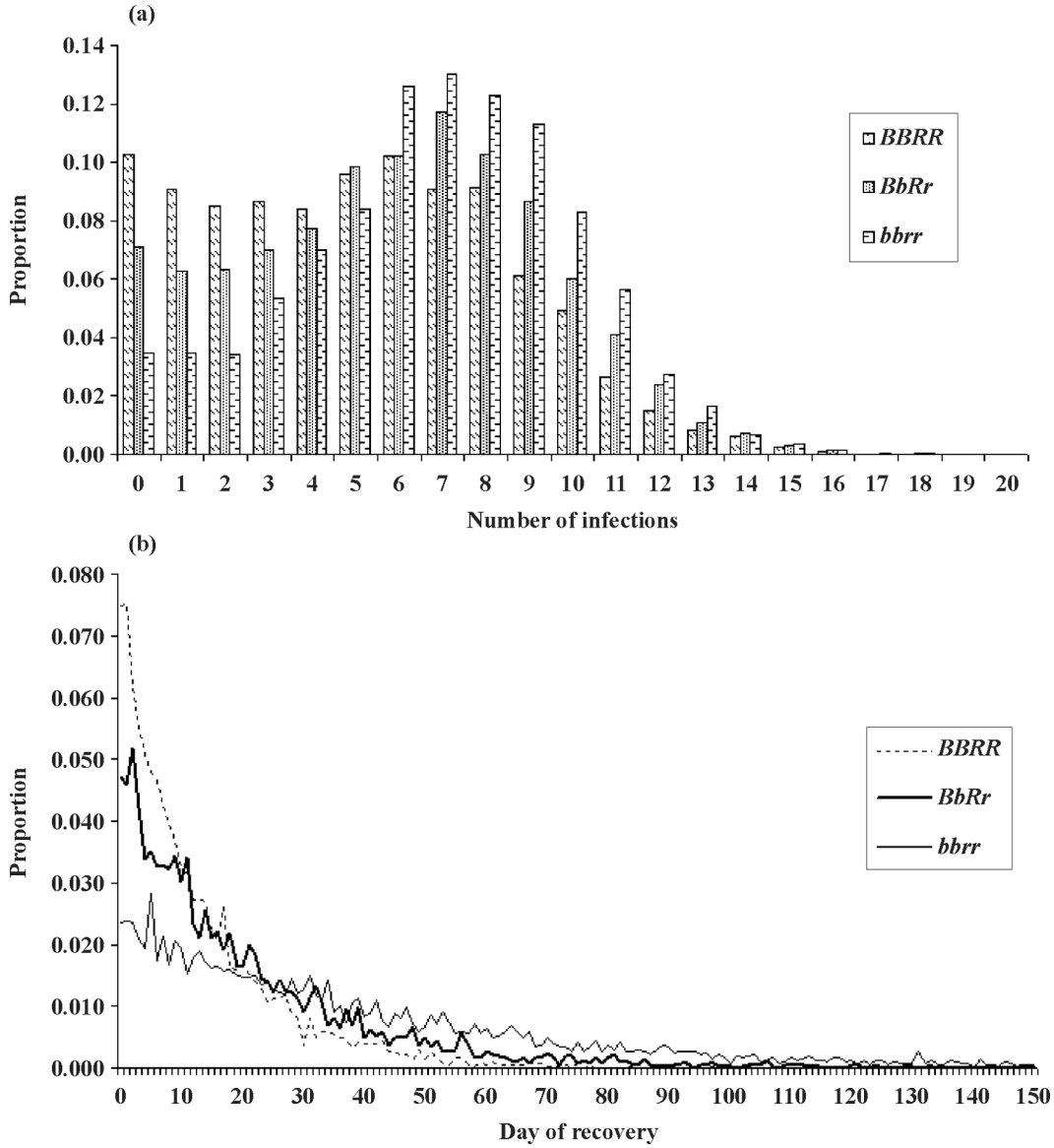

Figure 1. Frequency of proportions of (a) infections and (b) days of recovery of the index case of 3 genotypes when introduced into the population of 1,000 individuals with 3 genotypes. The alleles for the transmission coefficient (*B* and *b*) and recovery rate (*R* and *r*) are in Hardy-Weinberg equilibrium. Results are summarized from 5,000 replications.

pected by averaging expectations from homogeneous populations, hence slightly smaller inferred R_0 values. On the other hand, variation in transmission coefficients showed greater observed R_0 values than their theoretical expectations. However, it should be noted that in this paper, R_0 is estimated from the probability of no epidemic. Hence, the most robust interpretation is that it is the probability of no epidemic that is affected by heterogeneity, a result similar to that previously observed in a different context by Springbett et al. (2003). Thus, populations that are heterogeneous in host recovery (i.e., pathogen removal) pose a slightly lower epidemic risk than populations that are variable in susceptibility. Genetic heterogeneity also affected the time of occurrence of maximum epidemic severity, which was generally sooner than the expected value.

Under assumptions of no spatial heterogeneity and equal contact among animals of different genotypes, the prior expectation of the R_0 value in the heterogeneous population should be the weighted average of the R_0 values within each subgroup (Dushoff and Levin, 1995). A mathematical proof for a fully mixing genetically heterogeneous population comprising n genotypes was given by Bishop and MacKenzie (2003) using a simple *SIR* model. Similarly, in a differential susceptibility and infectivity model, Hyman and Li (2006) estimated the R_0 for the entire population as a weighted average of the R_0 of susceptible and infective subgroups. Hence, theoretical estimates may be considered as a first approximation to predict the general outcomes for a heterogeneous population that comprises 2 or more genotypes in different proportions; however,

specific epidemic properties (e.g., epidemic probabilities, severities, timing of maximum severity, etc.) may well differ from average values.

Some of the effects of (genetic) heterogeneity in both host and pathogen have previously been reported by Springbett et al. (2003), albeit in a slightly different context in which future disease risks as a function of heterogeneity per se were explored. The paper of Springbett et al. (2003) focused only on random variation in host susceptibility to infection, whereas in this paper, we consider the joint effects of defined variation in host susceptibility and recovery rate. One of our major aims was to explore the concepts of heterogeneity in a setting that may be more easily interpreted by animal geneticists, who seek to utilize disease resistance genes, than that presented by Springbett et al. (2003). Other studies have also considered effects of heterogeneity, albeit in different contexts from those explored here. For example, Yates et al. (2006) compared the risk of emergence of disease in homogeneous and heterogeneous populations, and they concluded that the influence of heterogeneity depends on the nature of the host variation as observed here. Other studies have suggested that increased heterogeneity leads to a de-

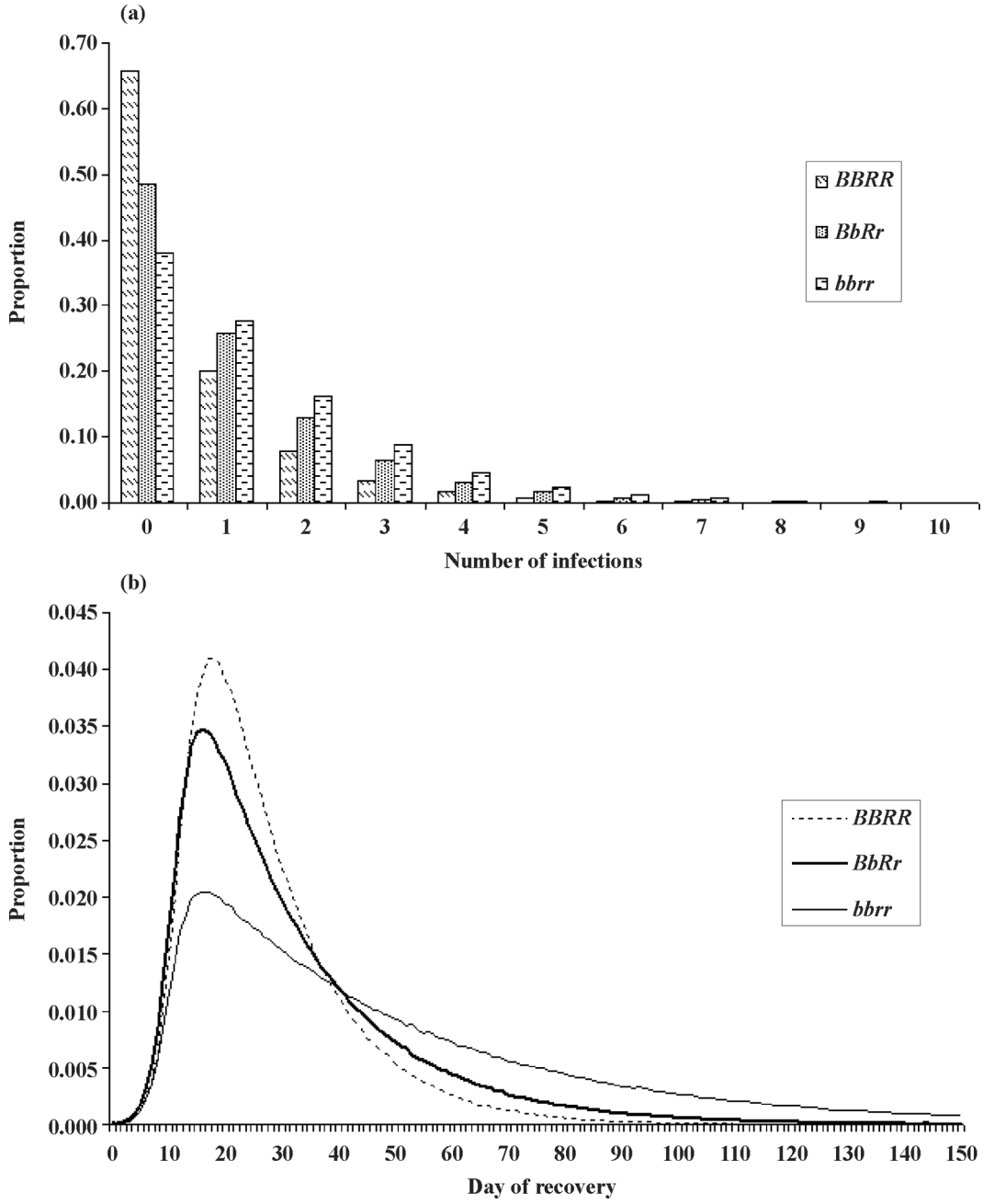

Figure 2. Frequency of proportions of (a) infections and (b) days of recovery of individuals of 3 genotypes observed in a population of 1,000 individuals with 3 genotypes. The alleles for the transmission coefficient (*B* and *b*) and recovery rate (*R* and *r*) are in Hardy-Weinberg equilibrium. The index case genotype is *BbRr*. Results are summarized from 5,000 replications.

crease in the probability of emergence of pathogens (Lloyd-Smith et al., 2005; Xiao et al., 2006).

Genetic Management for Disease Resistance

The parameterized model can be applied to identify disease management strategies that can exploit host genetic heterogeneity (e.g., to utilize animals that carry alleles for enhanced resistance to an infectious disease) and to identify strategies to minimize epidemic risks. An objective may be to construct a population containing sufficient resistant animals to reduce the expected R_0 to below 1.0, or at the least to reduce the probability of major epidemics to acceptable levels. Different strategies include selection and breeding of genetically resistant animals, constituting heterogeneous populations with different proportions of resistant and susceptible individuals, and avoiding contact between the infected and potentially susceptible individuals.

Evaluation of epidemic scenarios under selection and breeding of individuals of different genotypes is straightforward. For major genes, the proportion of offspring of different genotypes that would be produced from selected sires and dams of known genotypes is easily estimated. The model would then aid in identifying the risk of epidemic in a population that contains parents and offspring of diverse genotypes in variable proportions. Second, to manage the genetic structure of heterogeneously mixed populations, the present model can be explored to identify the proportion of individuals of each genotype that would constitute the final population so that the probability of occurrence of a major epidemic is reduced to an acceptable level, or so that R_0 goes below 1. This is more critical when there are limited numbers of resistant animals available or there are additional differences between different genotypes of animals. For example, resistant animals are sometimes less productive for important economic traits. Simple deterministic expectations in *SIR* models have been given by Bishop and MacKenzie (2003). However, the present model is advantageous over the simple deterministic expression, because it can consider host heterogeneity in more complex terms and gives the probability of epidemic outcomes. Among different nongenetic approaches, it may be of particular interest to restrict contacts between animals of different genotypes. A contact matrix corresponding to the contact rate (c_{ji}) between genotypes j and i may be formed, and possible epidemiological outcomes may be explored. For example, it is intuitively clear that limiting contact between infected individuals and noninfected susceptible individuals could decrease the disease incidence in the population. Further insights into the differential contact strategy may be drawn through introducing realistic farm structures and thereby creating different contact rates due to spatial heterogeneity. In general, the model is amenable to explore complex scenarios on genetic heterogeneity in which several strategies are combined together.

Effect of the Index Case

Another interesting feature in this study is the importance of the index case (i.e., the initial infected animal) in the overall epidemic. The index case may either recover, hence there is no epidemic, or it spreads the infection to other susceptible individuals and begins a potential epidemic. In the heterogeneous population investigated, the epidemic probabilities changed dramatically according to the recovery rate genotype of the index case, but subsequent epidemic events (like the maximum proportion infected or time of the maximum proportion of infection) were independent of index case genotype. As expected, index case genotypes with lower recovery rates remained in the population for a prolonged period of time, continuing to infect susceptible individuals and resulting in an increased chance of a catastrophic epidemic. This indicates that under practical situations, broad conclusions about epidemic risks will depend upon the genotype of the first infected animal. Because the transmission coefficient was modeled as a function of the susceptibility of the susceptible animal, the index case transmission coefficient genotype had no effect in our simulations. However, an equally valid parameterization would be to make the transmission coefficient a function of the infectivity of the infected animal. In this case, the index case genotype would affect subsequent epidemic outcomes. From a practical viewpoint, these results suggest that particular care should be taken to protect genotypes of high potential infectivity or low recovery rates, because if they were to become index cases, they would be more likely to cause severe epidemics than animals of less risky genotypes.

Purview of the Model

The model, as presented, is generalized for application to animal diseases, and as such it does have some limiting assumptions. First, it is important that a model should be parameterized in a way that is relevant to the disease of interest. An outline of the parameterization process was discussed by Nath et al. (2004). However, the underlying immunogenetic mechanisms controlling the host-pathogen interaction are not clearly understood for most diseases; hence, relating disease mechanisms to the appropriate compartmental epidemiological model could be difficult or intractable in some instances. Also, some aspects of disease biology are currently not in the model. For example, the present compartmental model does not consider individuals who have carrier status and continue to infect susceptible individuals without moving to the stage of recovery. Also, we have not included disease-dependent culling or disease-independent mortality in the present model. These aspects could be readily added to the current model.

On the other hand, the model can be easily generalized and hence is applicable to wider circumstances.

For example, it would be easy to extend the genetic heterogeneity as described in this paper to multiple loci, as well as integrate complex genetic relationships between alleles of different loci (for example, different signs and strengths of intra- and interallelic interaction). Similarly, the model can be updated to consider genetic resistance to be controlled by many genes, such as in the infinitesimal model. The utility can also be enhanced further by linking the current model with models that describe pathogen evolution under the effect of host heterogeneity.

To summarize, we have developed stochastic epidemic models that allow us to explore the effects of both mean genotype and genetic heterogeneity and challenge the assumptions that transmission of infection and epidemic severity is similar in homogeneous and heterogeneous fully mixing populations. We have demonstrated that genetic heterogeneity can affect the probabilities of epidemics, the severities of epidemics, and R_0, as derived from the probability of no epidemic, although the effects are subtle. Furthermore, we have shown that under some circumstances, epidemic outcomes depend upon the genotype of the index case. These results are of practical importance when determining how to best utilize disease resistant animals to minimize epidemic risks. The developed model can also be generalized to many complex situations related to host heterogeneity.

LITERATURE CITED

Adler, F. R. 1992. The effects of averaging on the basic reproduction ratio. Math. Biosci. 111:89–98.

Anderson, R. M., and R. M. May. 1992. Infectious Diseases of Humans: Dynamics and Control. 1st ed. Oxford Univ. Press, Oxford, UK.

Axford, R. F. E., S. C. Bishop, F. W. Nicholas, and J. B. Owen. 2000. Breeding for Disease Resistance in Farm Animals. 2nd ed. CABI, Oxon, UK.

Bishop, S. C. 2005. Disease resistance: Genetics. Pages 288–290 in Encyclopedia of Animal Science. W. G. Pond and A. W. Bell, ed. Marcel Dekker Inc., New York, NY.

Bishop, S. C., and K. MacKenzie. 2003. Genetic management strategies for controlling infectious disease in livestock populations. Genet. Sel. Evol. 35:S1–S17.

Bouma, A., M. C. M. De Jong, and T. G. Kimman. 1995. Transmission of pseudorabies virus within pig populations is independent of the size of the population. Prev. Vet. Med. 23:163–172.

Cheng, H. H. 2005. Integrated genomic approaches to understanding resistance to Marek's disease. Pages 24–33 in Proc. 3rd Int. Symp. Genet. Anim. Health. Iowa State Univ., Ames.

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1990. On the definition and the computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28:365–382.

Dushoff, J., and S. Levin. 1995. The effects of population heterogeneity on disease invasion. Math. Biosci. 128:25–40.

Hone, J. 1994. A mathematical model of detection and dynamics of porcine transmissible gastroenteritis. Epidemiol. Infect. 113:187–197.

Hyman, J. M., and J. Li. 2006. Differential susceptibility and infectivity epidemic models. Math. Biosci. Eng. 3:89–100.

Innocent, G. T., I. Morrison, J. Brownlie, and G. Gettinby. 1997. The use of a mass-action model to validate the output from a stochastic simulation model of bovine viral diarrhoea virus spread in a closed dairy herd. Prev. Vet. Med. 31:199–209.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438:355–359.

MacKenzie, K., and S. C. Bishop. 2001. Developing stochastic epidemiological models to quantify the dynamics of infectious diseases in domestic livestock. J. Anim. Sci. 79:2047–2056.

Nath, M., J. A. Woolliams, and S. C. Bishop. 2004. Identifying critical parameters in the dynamics and control of microparasite infection using a stochastic epidemiological model. J. Anim. Sci. 82:384–396.

Renshaw, E. 1991. Modelling Biological Populations in Space and Time. 1st ed. Cambridge Univ. Press, Cambridge, UK.

Springbett, A. J., K. MacKenzie, J. A. Woolliams, and S. C. Bishop. 2003. The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165:1465–1474.

Stark, K. D. C., D. U. Pfeiffer, and R. D. Morris. 2000. Within-farm spread of classical swine fever virus: A blueprint for a stochastic simulation model. Vet. Q. 22:36–43.

Xiao, Y., D. Clancy, N. P. French, and R. G. Bowers. 2006. A semi-stochastic model for *Salmonella* infection in a multi-group herd. Math. Biosci. 200:214–233.

Yates, A., R. Antia, and R. R. Regoes. 2006. How do pathogen evolution and host heterogeneity interact in disease emergence? Proc. R. Soc. Lond. B. Biol. Sci. 273:3075–3083.